

CLAIMS:

I claim:

1. A method for developing a script to be used with speech recognition systems, said method comprising the steps of:

reading language phoneme data for a given language, the language phoneme data having a plurality of phonemes occurring in the given language;

reading script data having a set of one or more phonemes;

counting each phoneme in the script data to produce a count data for each of the plurality of phonemes in the language phoneme data;

generating a set of statistical data derived from the count data, the set of statistical data including one or more metrics of the extent to which the phonemes in the language phoneme data are included in the script data.

2. The method of claim 1, wherein the script data includes one or more words, each word having one or more of the set of one or more phonemes, and further comprising:

reading vocabulary data having one or more words;

comparing each word in the script data with the vocabulary data; and

returning an error message if a word in the script data is not included in the vocabulary data.

3. The method of claim 2, wherein the step of counting each phoneme in the script data to produce a count data for each of the plurality of phonemes in the language phoneme data includes the steps of:

comparing each word in the script data with the vocabulary data;

returning an error message if a word in the script data is not included in the vocabulary data; and

counting each phoneme in each word in the script data if a word in the script data is included in the vocabulary data.

4. The method of claim 1, wherein the set of statistical data includes:

an occurrence data for each of the phonemes in the phoneme data, each occurrence data indicating a number of occurrences of the phoneme in the script data.

5. The method of claim 1, wherein the set of statistical data includes:

a ratio data, each ratio data being the number of phonemes in the script data as a percentage of the number of the plurality of phonemes in the phoneme data.

6. The method of claim 1, wherein the set of statistical data includes:

a missing phoneme data, each missing phoneme data being a list of the phonemes in the language phoneme data not included in the script data.

7. The method of claim 1, wherein the script data includes one or more words, and further comprising the steps of:

- reading a vocabulary data having one or more words;
- reading an additional word having one or more phonemes;
- comparing the additional word with the vocabulary data;
- adding the additional word to the script data if the additional word is included in the vocabulary data.

8. The method of claim 1, wherein the script data includes one or more words, and further comprising the steps of:

- reading a vocabulary data having one or more words;
- reading an additional word having one or more phonemes;
- comparing the additional word with the script data;
- removing the additional word from the script data if the additional word is included in the script data.

9. The method of claim 1, wherein the script data includes one or more words, and further comprising the steps of:

reading a vocabulary data having one or more words;

reading a set of one or more desired phonemes;

searching the vocabulary data for one or more words having the set of one or more desired phonemes;

generating a report of one or more additional words having the set of one or more desired phonemes, if the one or more additional words having the set of one or more desired phonemes are included in the vocabulary data.

10. A machine readable storage having stored thereon a computer program for developing a script to be used with speech recognition systems, said computer program comprising a routine set of instructions for causing the machine to perform the steps of:

reading a language phoneme data for a given language, the language phoneme data having a plurality of phonemes occurring in the given language;

reading a script data having a set of one or more phonemes;

counting each phoneme in the script data to produce a count data for each of the plurality of phonemes in the language phoneme data;

generating a set of statistical data derived from the count data, the set of statistical data including one or more metrics of the extent to which the phonemes in the language phoneme data are included in the script data.

11. The machine readable storage of claim 10, wherein the script data includes one or more words, each word having one or more of the set of one or more phonemes, and for further causing said machine to perform the steps of:

- reading a vocabulary data having one or more words;
- comparing each word in the script data with the vocabulary data; and
- returning an error message if a word in the script data is not included in the vocabulary data.

12. The machine readable storage of claim 11, wherein the step of counting each phoneme in the script data to produce a count data for each of the plurality of phonemes in the language phoneme data includes the steps of:

- comparing each word in the script data with the vocabulary data;
- returning an error message if a word in the script data is not included in the vocabulary data; and
- counting each phoneme in each word in the script data if a word in the script data is included in the vocabulary data.

13. The machine readable storage of claim 10, wherein the set of statistical data includes:

- an occurrence data for each of the phonemes in the phoneme data, each occurrence data indicating a number of occurrences of the phoneme in the script data.

14. The machine readable storage of claim 10, wherein the set of statistical data includes:

a ratio data, each ratio data being the number of phonemes in the script data as a percentage of the number of the plurality of phonemes in the phoneme data.

15. The machine readable storage of claim 10, wherein the set of statistical data includes:

a missing phoneme data, each missing phoneme data being a list of the phonemes in the language phoneme data not included in the script data.

16. The machine readable storage of claim 10, wherein the script data includes one or more words, and further causing the machine to perform the steps of:

reading a vocabulary data having one or more words;

reading an additional word having one or more phonemes;

comparing the additional word with the vocabulary data;

adding the additional word to the script data if the additional word is included in the vocabulary data.

17. The machine readable storage of claim 10, wherein the script data includes one or more words, and further causing the machine to perform the steps of:

reading a vocabulary data having one or more words;

reading an additional word having one or more phonemes;

comparing the additional word with the script data;

removing the additional word from the script data if the additional word is included in the script data.

18. The machine readable storage of claim 10, wherein the script data includes one or more words, and further causing the machine to perform the steps of:

reading a vocabulary data having one or more words;

reading a set of one or more desired phonemes;

searching the vocabulary data for one or more words having the set of one or more desired phonemes;

generating a report of one or more additional words having the set of one or more desired phonemes, if the one or more additional words having the set of one or more desired phonemes are included in the vocabulary data.

19. A script development tool configured for coupling to a script having a set of one or more phonemes and programmed to both count each phoneme in said script to produce count data for each phoneme in a selected language, and also to generate a set of statistical data derived from said count data, the set of statistical data comprising one or more metrics of the extent to which each phoneme in said selected language is included in said script.

20. The tool of claim 19, wherein the script includes one or more words, and wherein the tool is further programmed to read a vocabulary data having one or more words, and to read an additional word having one or more phonemes, and is also programmed to compare the additional word with the vocabulary data and add the additional word to the script data if the additional word is included in the vocabulary data, and is also programmed to compare the additional word with the script and remove the additional word from the script data if the additional word is included in the script data.